

# REFINING AGRICULTURAL OFFICIAL STATISTICS THROUGH MIXED METHODS: THE INTEGRATION OF AREA SAMPLING FRAME AND REMOTE SENSING

Arif Handoyo Marsuhandi, Achmad Fauzi Bagus Firmansyah, Nasiya Alifah Utami, Novia Permatasari, Maria Shawna Cinnamon Claire, Wida Widiastuti, Fawcet Jenusdy Makay, Aldi Saeful Rauf, Casslirais Surawan, Sugiri, Aldo Leofiro Irwansyah, Yudhi Agustar Sanjaya, Muhammad Zabbar Falihin, Alfatihah Reno Maulani Nuryaningsih Soekri Putri Munaf, Setia Pramana, Kadarmento

Statistics Indonesia, Jakarta, Indonesia - [arifhandoyo@bps.go.id](mailto:arifhandoyo@bps.go.id)

As the National Statistics Office (NSO) of Indonesia, Statistics Indonesia (BPS) has taken a strategic step toward integrating big data into the production of agricultural official statistics. While big data has traditionally been used as a complementary analytical resource within national statistical systems, this study advances its role toward becoming an operationally integrated component of official statistical production. The research develops and evaluates a Mixed Method framework that integrates Area Sampling Frame (ASF) survey data, Sentinel-1 Synthetic Aperture Radar (SAR) imagery, and machine learning techniques to improve the timeliness, spatial granularity, and reliability of rice phenology statistics. Within this framework, ASF survey data function as ground-truth references, while satellite-derived backscatter variables extend spatial coverage beyond sampled locations. The modelling stage applies the XGBoost algorithm, followed by a Bayesian post-processing procedure that incorporates temporal transition probabilities to enhance logical consistency in growth-stage predictions. Under controlled modelling conditions, the approach achieved high classification accuracy (91-97%), with Bayesian correction significantly improving temporal stability. When implemented nationwide across all KSA sample points without filtering, performance decreased due to increased heterogeneity but remained substantially improved after post-processing (accuracy increasing from 0.595 to 0.752), demonstrating robustness under real-world conditions. Spatial evaluation revealed regional variability, reflecting differences in agricultural complexity and cropping systems. Importantly, the framework is designed in alignment with the Fundamental Principles of Official Statistics, ensuring transparency, methodological rigor, and institutional accountability. The results demonstrate that integrating survey data, remote sensing, and probabilistic machine learning provides a feasible and governance-consistent pathway for modernizing agricultural statistics, strengthening evidence-based policymaking, enhancing food security monitoring, and improving statistical coverage across Indonesia's geographically diverse archipelago.

**Keywords:** Area Sampling Frame; Remote Sensing; Official Statistics; Machine Learning.

The authors declare that the manuscript is original and that all text and materials included in this work are free from copyright violations and properly cited where applicable.

## 1. Introduction

Agriculture remains one of the most important sectors in Indonesia's economy, contributing significantly to regional gross domestic product and employing nearly one-third of the national labor [1][2]. The strategic role of agriculture underscores the critical need for reliable and timely agricultural statistics to support evidence-based policymaking and food security planning. Improving the quality of official agricultural statistics has therefore become a priority for modern national statistical systems.

To address this challenge, Statistics Indonesia (BPS) introduced an Area Sampling Frame (ASF) in 2018 to replace conventional eye-estimate methods for measuring harvested area. The ASF represents a major methodological advancement by incorporating geospatial information and satellite imagery into survey design, consistent with recommendations from the Food and Agriculture Organization (FAO) [3][4]. While this framework has improved objectivity and accuracy, operational challenges remain, particularly in remote and high-risk areas where field data collection is costly, difficult, and sometimes unsafe.

Recent advances in remote sensing and machine learning offer new opportunities to complement field-based surveys and enhance the production of agricultural statistics. Although numerous studies have demonstrated the potential of these technologies for crop monitoring and yield estimation [5][6][7], their systematic integration into official statistical systems remains limited. Existing applications are majority to research settings or pilot projects and lack an operational framework that satisfies the Fundamental Principles of Official Statistics. Failure to comply with these principles may expose National Statistical Offices (NSOs) to certain risks [8].

This study addresses this gap by proposing and evaluating an integrated mixed-method framework that combines ASF-based ground observations, satellite remote sensing, and machine learning techniques for the production of agricultural official statistics. The proposed framework leverages existing survey infrastructure while incorporating advanced modeling to generate timeliness, spatially granular, and statistically robust estimates. The study assesses the effectiveness of this integration in overcoming logistical constraints, improving spatial resolution, and maintaining statistical reliability within Indonesia's official statistical system. By bridging methodological innovation with operational requirements, this research contributes a practical framework for modernizing agricultural statistics in large and geographically diverse countries.

## 2. Analytical Framework

This study is based on an analytical framework designed to support the systematic integration of official statistical sources with emerging satellite-based data to advance the modernization of statistical governance. The framework combines ASF survey data, an officially validated and integral component of the current official statistical system, with satellite imagery, which provides enhanced spatial and temporal coverage. Within this structure, the ASF dataset functions as ground-truth reference information for model training and calibration, enabling the extension of survey-based statistics through their integration with satellite observations, in line with established practices of combining survey and Earth observation datasets in official statistical systems [9].

The framework focuses on the coordinated use of traditional and innovative data sources to strengthen the efficiency, relevance, and long-term sustainability of official statistical production. Satellite and geospatial data recognized by national statistical organizations as valuable complementary sources that address gaps in timeliness, coverage, and spatial detail where traditional survey and administrative sources alone may be

inadequate [10]. ASF survey data provide historically consistent and quality-assured measurements since 2018, while satellite data supply continuous, large-scale observations that mitigate limitations related to coverage and timeliness. By linking these sources within a unified analytical structure, the model translates statistical information into spatially explicit indicators suitable for policy analysis, monitoring, and evaluation, enhancing the utility of official statistics for evidence-based policymaking [11].

From a governance perspective, this hybrid framework contributes to the modernization of national statistical systems by leveraging existing institutional infrastructures while incorporating emerging technological capabilities. The integration of survey data with geospatial information enhances interoperability, transparency, and methodological reproducibility within the statistical production process. This approach is consistent with international guidance emphasizing that Earth observation data “can complement traditional data sources” and should be integrated into “official statistical production processes” to improve timeliness, coverage, and policy relevance [10]. Furthermore, the importance of “mainstreaming geospatial information into official statistics” to strengthen national statistical systems and support sustainable development monitoring has been highlighted in global statistical recommendations [12]. By embedding geospatial integration within established survey frameworks, the proposed model aligns with principles of institutional coordination, transparency, and responsible statistical modernization [10][12]. This approach strengthens institutional capacity to deliver timely, policy-relevant evidence, promotes coherence within the official statistics ecosystem, and supports informed and adaptive decision-making while maintaining continuity with established statistical standards and practices.

### 3. Methodology

#### Workflow of Mixed Method

To achieve the objectives of this study, a modelling workflow is proposed, as illustrated in Figure 1. This workflow follows a standard structure consisting of three main stages: pre-modelling, modelling, and post-processing.

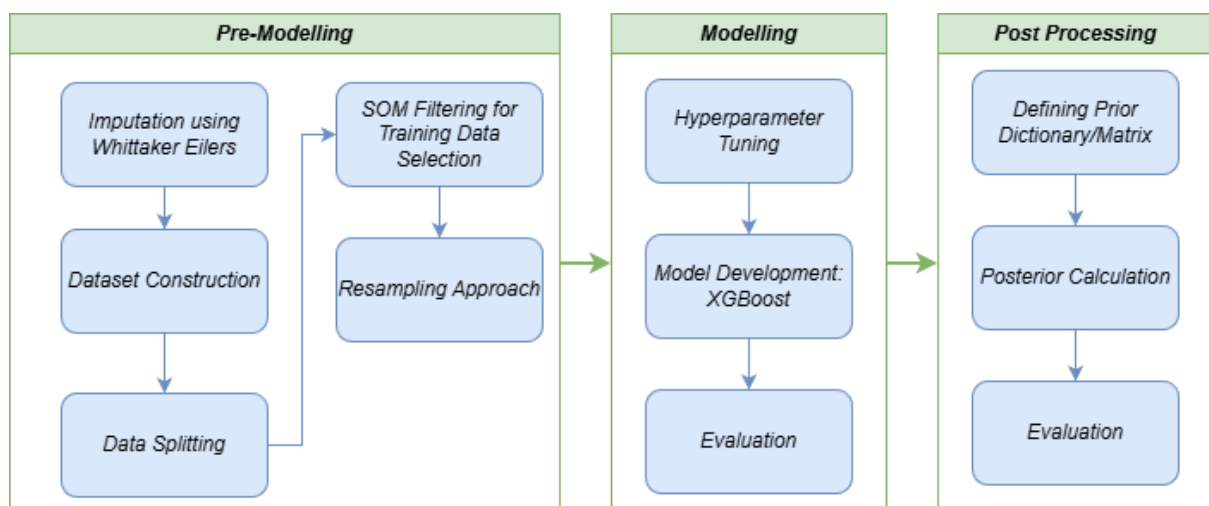


Figure 1. Workflow of Mixed Method (Source: Mixed Method Team, BPS)

In the pre-modelling stage, the ASF survey dataset is integrated with Sentinel-1 satellite imagery after conducting data cleaning to ensure completeness and consistency of the survey

data and applying standardized preprocessing procedures to generate calibrated satellite-derived variables. The integrated dataset is then examined to verify its suitability for analysis, followed by filtering to remove irrelevant or low-quality observations.

Additional preprocessing steps are applied to address potential class imbalance, as imbalanced data can negatively affect model performance and increase the risk of overfitting or underfitting; therefore, balancing techniques are implemented to enhance model generalizability. The workflow then proceeds to the modelling stage, where the XGBoost algorithm is trained and optimal hyperparameters are determined through systematic tuning to achieve stable and reliable predictive performance, accompanied by preliminary evaluation to assess model behavior. Finally, the post-processing stage evaluates and refines the modelling results to ensure robustness and logical consistency, validating that the final estimates are accurate, reliable, and suitable for supporting statistical analysis and evidence-based decision-making within the official statistics framework.

### **Area Sampling Frame**

To enhance the accuracy and reliability of rice production statistics, Statistics Indonesia (BPS) in collaboration with the National Research and Innovation Agency (BRIN) developed the Area Sampling Frame (ASF) as an improved methodology for estimating harvested rice area. Prior to the adoption of ASF, harvested area data were primarily derived from visual eye-estimate approaches, which are prone to subjectivity and measurement bias. The ASF is a probability-based, area-oriented sampling framework that uses geographically referenced and permanently fixed sample points systematically distributed across the study region. These points are observed in situ by trained enumerators following standardized field protocols. Since 2018, ASF has officially implemented the national standard method for collecting harvested rice area statistics throughout Indonesia.

### **Sentinel-1**

This study uses radar-based satellite imagery as its primary data source. The selection of radar imagery motivated by Indonesia's geographic characteristics as an archipelagic country where persistent cloud cover frequently limits the usability of optical satellite data. In alignment with the principles of official statistics particularly regarding data reliability, accessibility, and continuity imagery from the Sentinel-1 mission selected as the core input for the modeling framework.

Sentinel-1 is a C-band Synthetic Aperture Radar (SAR) mission under the Copernicus Programme, operated by the European Space Agency (ESA). It provides systematic Earth surface observations day and night and is largely independent of weather conditions. The Sentinel-1 sensor acquires data in dual polarization modes, namely VV (vertical-vertical) and VH (vertical-horizontal). VV polarization is primarily sensitive to surface roughness, whereas VH polarization is more responsive to volume scattering associated with vegetation structure. With a spatial resolution of approximately 10-20 meters, depending on the acquisition mode, the combined use of these polarizations enables detailed characterization of terrestrial and marine surfaces for applications such as agricultural monitoring, forest biomass assessment, land deformation analysis, and maritime surveillance.

In this study, Sentinel-1 imagery was acquired through collaboration with the National Research and Innovation Agency of Indonesia (BRIN). Prior to analysis, the data underwent standardized preprocessing procedures, including geometric correction and radiometric calibration, to ensure spatial accuracy and radiometric consistency. These preprocessing steps reduce potential distortions, noise, and systematic bias inherent in raw SAR data. As a result, the processed Sentinel-1 imagery provides more reliable and consistent input features,

thereby improving the accuracy, robustness, and overall quality of the subsequent analysis and modelling results.

### **Data Integration between ASF Survey and Satellite Imagery**

This spatial integration ensures that both datasets share a common spatial reference framework, enabling the direct association of satellite features with corresponding ASF ground observations. The resulting integrated dataset combines the statistical reliability of survey data with the spatial continuity of satellite imagery. To ensure data quality and consistency, additional filtering and validation procedures are applied after integration. Observations with incomplete information, spatial mismatch, or low-quality satellite signals are excluded from further analysis. This process produces a harmonized dataset suitable for statistical modelling and machine learning applications. The integrated dataset serves as the primary analytical foundation for model development, allowing satellite-derived variables to be used as predictors and ASF survey variables to serve as reference labels in the classification process.

### **Classification Method Using XGBoost**

Extreme Gradient Boosting (XGBoost) algorithms are the primary classification method in this study, leveraging its strong predictive performance, computational efficiency, and robustness in handling structured and heterogeneous datasets. Within this framework, ASF survey data serve as reference classification labels, while satellite-derived backscatter variables function as predictor features, allowing the model to learn the relationship between radar observations and corresponding phenological classes and to extend predictions beyond the original survey locations. Prior to training, preprocessing procedures are applied to address potential class imbalance, which may otherwise bias the model toward dominant classes and reduce accuracy for minority categories; appropriate balancing techniques are therefore implemented to enhance generalizability and reliability.

### **Temporal Imputation Using Whittaker-Eilers Smoothing**

Missing or irregular temporal observations in satellite image variables are reconstructed using the Whittaker-Eilers smoothing approach. This penalized least squares method is particularly suitable for remote sensing time-series data, as it preserves the underlying signal structure while reducing noise. The smoother minimizes:

$$Q = \sum_{i=1}^n w_i (y_i - z_i)^2 + \lambda \sum_{i=2}^n (\Delta^2 z_i)^2$$

where  $y_i$  represents observed values,  $z_i$  the smoothed estimates,  $w_i$  weights, and  $\lambda$  the smoothing parameter controlling the bias variance tradeoff. This procedure enhances temporal stability and reduces signal irregularities prior to modelling.

### **Training Data Preparation and Modelling Strategy**

To enhance data quality and model robustness, several preparatory and modelling procedures are implemented prior to final classification. First, training data filtering is conducted using a Self-Organizing Map (SOM) approach. SOM performs topology-preserving clustering in high-dimensional feature space, enabling the identification

of anomalous, sparse, or noisy observations. By filtering samples based on SOM cluster structure, the training dataset becomes more homogeneous and representative, thereby improving model stability and reducing the influence of outliers.

To address potential class imbalance, an adaptive resampling strategy is applied. For minority classes containing fewer than 1,000 samples, Borderline-SMOTE is used to generate synthetic observations near class decision boundaries, increasing sensitivity in ambiguous regions and improving class representation. The resampling process continues until each minority class reaches 1,000 samples. For majority classes exceeding 5,000 samples, the Neighbourhood Cleaning Rule (NCL) is implemented to remove noisy or ambiguous instances, refine class boundaries, and reduce bias caused by overrepresented categories. After balancing, the final dataset is partitioned into training and testing subsets to ensure objective evaluation and prevent information leakage.

The modelling process proceeds by training the classification model using the training subset, while the testing subset is reserved for performance evaluation. Hyperparameter tuning is conducted to identify the optimal model configuration, and cross-validation is applied to ensure consistent performance across different data partitions, thereby reducing the risk of overfitting.

This study uses the Extreme Gradient Boosting (XGBoost) algorithm as the primary classification model. XGBoost builds an ensemble of decision trees sequentially to improve predictive accuracy and is well suited for structured datasets with complex nonlinear relationships. The algorithm learns patterns by linking satellite-derived features, such as radar backscatter values, with known ASF survey classifications. Once trained, the model is capable of predicting rice growth stages in areas where only satellite data are available, thereby extending the spatial coverage of survey-based statistics within the integrated framework.

## **Post Modelling**

The post-modelling stage ensures that machine learning outputs are statistically valid, methodologically transparent, and institutionally aligned with the principles governing official statistics production. In this study, post-modelling procedures are designed not only to evaluate predictive performance but also to guarantee coherence with established ASF survey-based standards. Model outputs are first validated using an independent testing dataset derived from the ASF framework, where standard classification metrics including overall accuracy, precision, recall, and F1-score are computed to assess predictive performance. Beyond these metrics, validation emphasizes statistical reliability and consistency with the ground-truth survey system. The confusion matrix is carefully examined to identify systematic misclassification patterns, particularly errors that could introduce structural bias when results are aggregated at higher reporting levels.

To maintain consistency with official definitions, all classification outputs are reviewed against established ASF agricultural classification standards. Where necessary, probability thresholds are adjusted to ensure conceptual alignment between model predictions and official survey-based categories. This alignment safeguards comparability over time and preserves coherence within the national statistical system.

Finally, model-based predictions generated at fine spatial resolution (e.g., pixel or grid level) are aggregated to official reporting units such as districts or provinces. The aggregation process follows standard statistical compilation procedures to ensure compatibility with existing reporting frameworks. Comparisons with historical statistical series are conducted to detect structural deviations, and substantial discrepancies are further examined to determine whether they reflect genuine agricultural dynamics or modelling artefacts. Through these

procedures, the post-modelling stage ensures that model-based estimates remain reliable, statistically consistent, and operationally suitable for official dissemination.

## 4. Result

### Modelling Outcomes

The performance results are presented using two metrics: overall accuracy and F1-macro. Accuracy reflects the model's general performance across all classes, while F1-macro accounts for class imbalance by computing the harmonic mean of precision and recall for each class and averaging them equally. This ensures that minority classes contribute proportionally to the final performance score. The accuracy outcomes are visualized in Figure 2, which maps the accuracy values across provincial boundaries. For interpretability, the accuracy levels are categorized into three groups:

- Less than 70%, indicating lower model performance;
- 70-85%, indicating moderate performance; and
- Greater than 85%, indicating high performance.

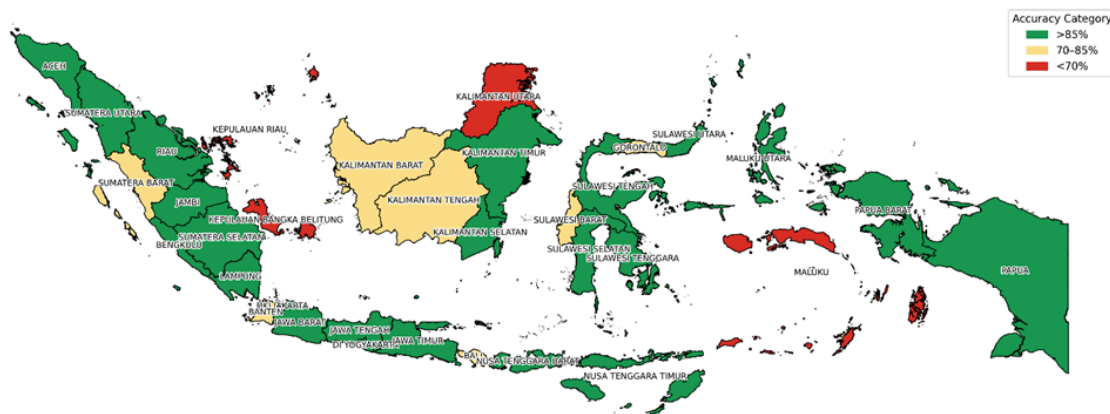


Figure 2. Map of accuracy distribution by Province in Indonesia 2025

Based on the spatial distribution, most provinces fall into the high-performance category. Notably, all rice self-sufficient provinces such as East Java, Central Java, West Java, South Sumatra, Lampung, North Sumatra, and South Sulawesi are included in this group. These provinces have extensive paddy field areas, which provide a larger volume of training data and consequently contribute to stronger model performance.

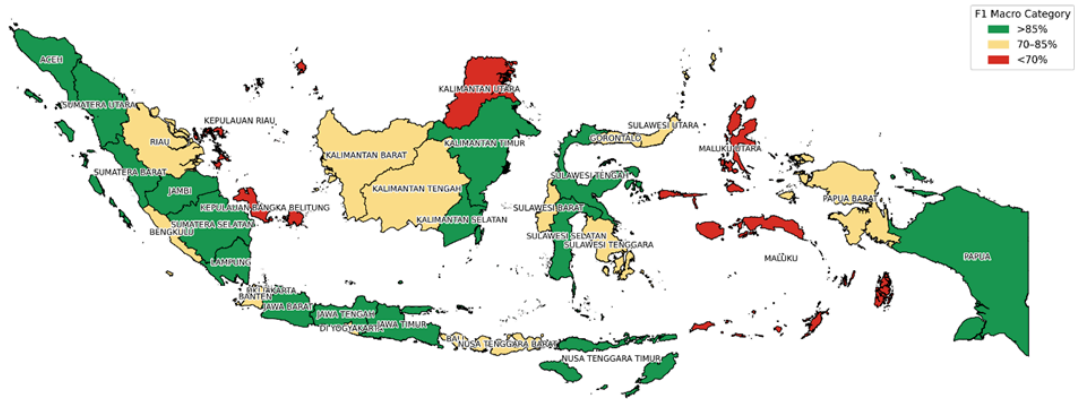


Figure 3. Map of Model F1 Macro distribution by Province in Indonesia 2022

Figure 3 presents the model performance using the F1-macro metric, which explicitly accounts for class imbalance in the dataset. Overall, several provinces remain within the same performance categories as observed in the accuracy-based assessment. However, notable shifts are observed in a number of provinces. Specifically, West Sumatra and Bengkulu move from the moderate-performance category to the high-performance category when evaluated using F1-macro. In contrast, Riau, the Special Region of Yogyakarta (DIY), West Nusa Tenggara, North Sulawesi, Southeast Sulawesi, and West Papua shifted from the high-performance category to the moderate-performance category. Additionally, Maluku experiences a more pronounced decline, moving from the high-performance category to the low-performance category.

As a follow-up step, the model was then applied to make predictions for all available data, totaling 25,095,809 points from the field activities of the ASF Survey covering all regions of Indonesia. This application produced a confusion matrix that provides a comprehensive overview of the model's ability to classify the data accurately, while also reinforcing the findings related to performance variations across provinces.

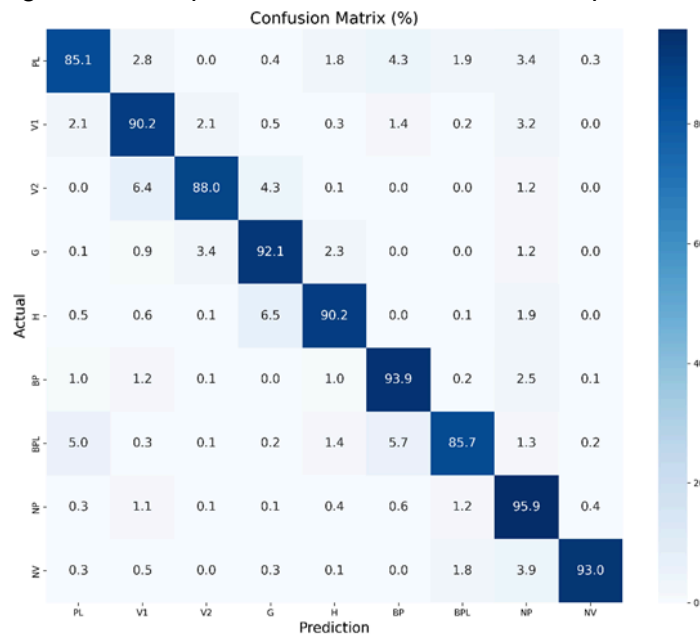


Figure 4. Confusion Matrix of Mixed Method

Figure 4 presents the confusion matrix summarizing the model's performance in classifying rice phenology at the national scale. Overall, the model exhibits strong performance across all phenological classes as well as the additional categories (Non-Paddy and Non-Vegetation). Several classes achieve notably high accuracy, such as Non-Paddy (NP)

with 95.9%, Fallow-After-Harvest (BP) with 93.9%, and Non-Vegetation (NV) with 93.0%. The primary stages of rice phenology also demonstrate robust performance, including the Generative (G) stage at 92.1%, Harvest (H) at 90.2%, and Vegetative 1 (V1) also at 90.2%. On the other hand, the Tillage or Land Preparation (PL) and Fallow-During-Tillage (BPL) stages show slightly lower accuracies 85.1% and 85.7% respectively yet both remain within the “good” performance range, as indicated by values above 85%.

### Post-Processing Results

The post-processing results are evaluated using accuracy-based performance metrics to assess the refinement of machine learning predictions within a logically structured modelling framework. Using the same dataset and visualization scheme as in the modelling stage to ensure analytical consistency, the comparison between the raw XGBoost output and the Bayesian-corrected results demonstrates a substantial improvement in classification performance. The initial XGBoost model achieved an overall accuracy of 91.44%, but the confusion matrix revealed notable misclassifications, particularly between the Fallow-After Harvest (BP) and Fallow-During Tillage (BPL) phases, as well as between Early and Late Vegetative stages (V1 and V2), reflecting spectral overlap commonly observed in radar-based phenological analysis. After applying Bayesian post-processing that incorporates temporal prior information to enforce agronomically plausible growth transitions, overall accuracy increased significantly to 97.48% (Figure 5).

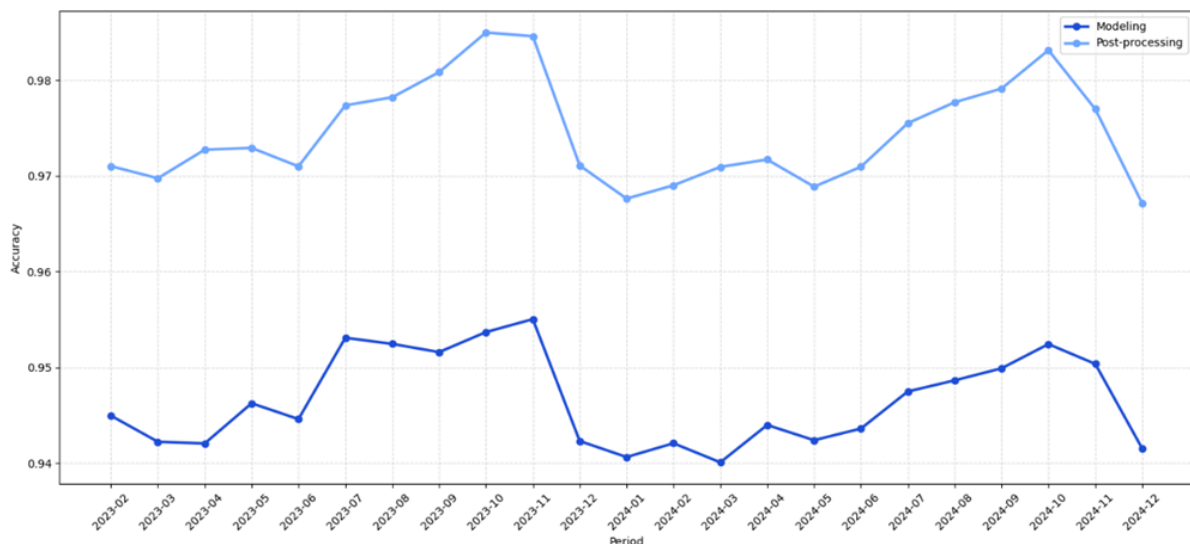


Figure 5. Modelling Accuracy VS Post-processing Accuracy

The improvement is reflected in a stronger dominance of diagonal elements in the confusion matrix and marked gains across nearly all classes, especially fallow and vegetative phases. Monthly accuracy analysis further confirms that post-processing enhances temporal stability, with accuracies consistently exceeding 96% and reaching above 98% during several periods, while also mitigating seasonal fluctuations observed in the raw model. These findings demonstrate that integrating XGBoost classification with Bayesian temporal correction effectively reduces spectral ambiguity and temporal inconsistencies, thereby improving the reliability and stability of rice phenology identification using Sentinel-1 imagery.

### Application of the Mixed Method Model and Algorithm Across Indonesia

After modelling and post-modelling, The Mixed Method model and Bayesian post-processing algorithm were subsequently applied to all KSA sample points across Indonesia. When implemented nationwide, the raw XGBoost model achieved an overall accuracy of 0.595,

indicating substantial inter-class misclassification. However, figure 6 shows after applying Bayesian temporal correction, accuracy improved significantly to 0.752, with stronger diagonal dominance in the confusion matrix and notable gains in classes such as G, H, V2, and NV. Although this performance remains lower than the 91%-97% accuracy obtained during the controlled modelling phase using filtered data, the decline is expected given the increased complexity, spectral noise, spatial variability, and shifting sample conditions present in nationwide application.

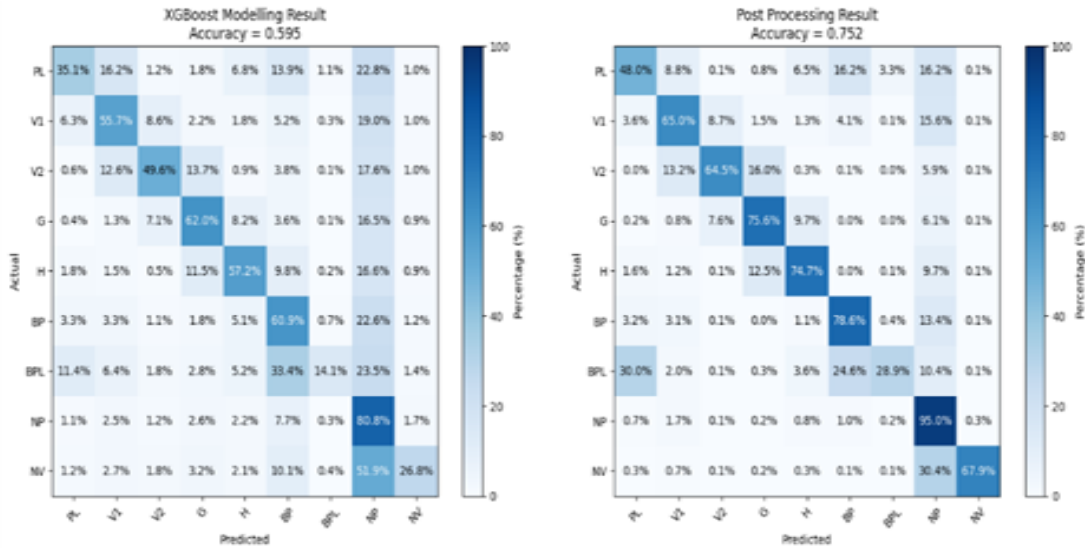


Figure 6. Modelling without and with Post-processing

Spatial evaluation further in figure 7 shows that model performance is generally moderate to high ( $\geq 70\%$ ) across most provinces, with the strongest results observed in Papua, Maluku, and North Maluku ( $>85\%$ ), likely due to more homogeneous agricultural patterns or dominance of non-rice classes,

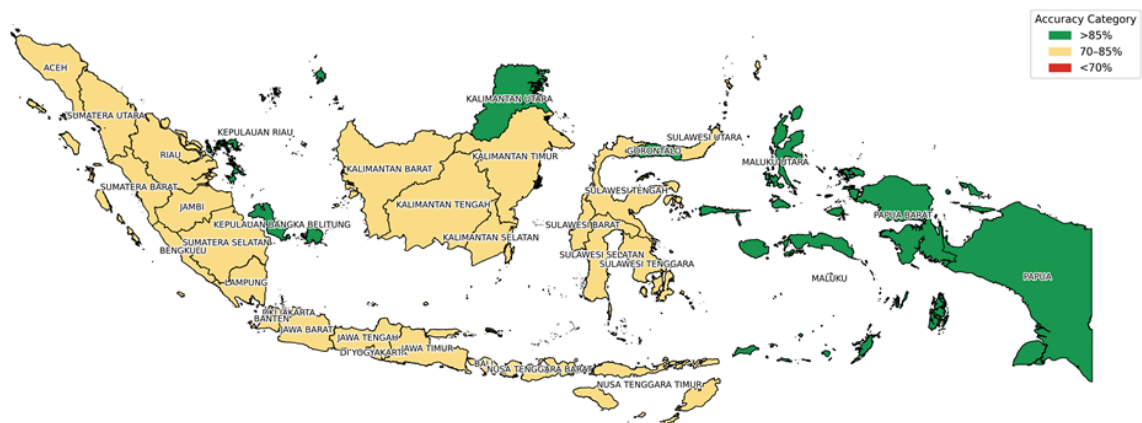


Figure 7. Accuracy Map Across of Indonesia

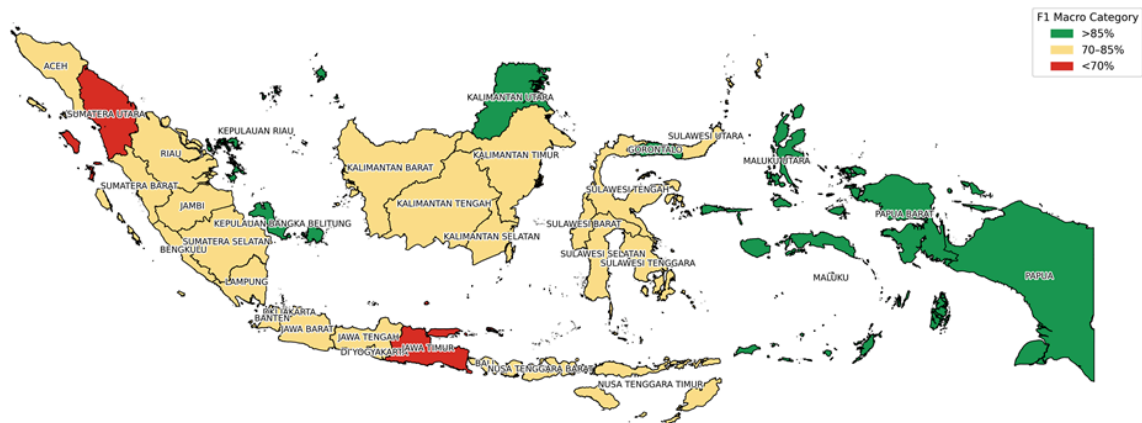


Figure 8. F1-Macro Map Across of Indonesia

While in Figure 8, relatively lower F1-macro values in parts of Java and North Sumatra reflect intensive and complex cropping systems that challenge phenological separability. District-level analysis confirms substantial regional variation, with lower accuracies concentrated in central-southern Sumatra, much of Java, and parts of NTT and South Kalimantan areas characterized by heterogeneous land conditions and unstable spectral signals whereas higher accuracies in Papua, Maluku, parts of Kalimantan, central-northern Sulawesi, and selected provinces such as South Sulawesi, West Java, and West Sumatra suggest more consistent phenological patterns that are easier for the model to capture. These findings demonstrate that Bayesian post-processing substantially enhances model stability and robustness under real-world conditions, while also highlighting that regional agricultural complexity remains a key determinant of predictive performance and a priority area for further methodological refinement.

## 5. Conclusion

This study presents a framework for refining agricultural official statistics through mixed methods, as reflected in the integration of Area Sampling Frame (ASF) survey data and remote sensing within Indonesia's official statistical system. By positioning ASF data as ground-truth references and leveraging Sentinel-1 SAR imagery to enhance spatial and temporal coverage, the study demonstrates how traditional survey infrastructure and Earth observation data can be systematically combined to strengthen statistical production. The structured workflow comprising pre-modelling, modelling, and post-modelling stages ensured data harmonization, class balancing, and rigorous validation. Under controlled conditions, the XGBoost-based classification achieved high accuracy (91-97%), while Bayesian post-processing significantly improved temporal coherence and increased accuracy from 91.44% to 97.48% by reducing misclassification among spectrally similar rice growth stages.

When operationalized nationwide without data filtering, model performance declined due to heterogeneous agricultural conditions and real-world spectral variability; however, Bayesian refinement improved accuracy from 0.595 to 0.752, confirming the robustness of the mixed-method approach under practical implementation. Spatial analysis further highlighted regional performance differences linked to cropping complexity and land-use diversity. Overall, the findings demonstrate that integrating ASF surveys with remote sensing and probabilistic modelling provides a feasible, scalable, and governance-aligned pathway for modernizing agricultural official statistics. This mixed-method integration not only enhances spatial granularity and timeliness but also maintains consistency with established statistical standards, thereby supporting evidence-based agricultural planning, food security monitoring, and sustainable policy development in Indonesia.

## References

1. Badan Pusat Statistik (BPS). PDRB Indonesia: tabel dinamis [Internet]. Jakarta: BPS; [cited 2026 Feb 20]. Available from: <https://www.bps.go.id>
2. Badan Pusat Statistik (BPS). Potensi pertanian Indonesia: peta baru pertanian berkelanjutan. Jakarta: BPS; 2024.
3. Badan Pusat Statistik (BPS). Upaya perbaikan data padi dengan metode kerangka sampel area (KSA) 2018. Jakarta: BPS; 2018.
4. Food and Agriculture Organization of the United Nations (FAO). Handbook on remote sensing. Rome: FAO; 2015.
5. Marsuhandi AH, Soleh AM, Wijayanto H, Domiri DD. Utilization of ensemble learning and remote sensing for rice land type classification. In: *Proc Semin Nas Off Stat 2019: Development of Official Statistics for Supporting SDGs Implementation*; 2019; Jakarta, Indonesia. Jakarta: Politeknik Statistika STIS; 2019. p.188-195.
6. Triscowati DW, Buana WP, Marsuhandi AH. Mapping corn land potential using satellite imagery and Random Forest on Google Earth Engine cloud computing. In: *Proc Semin Nas Off Stat 2021: Official Statistics Development for Supporting SDGs Implementation*; 2021 Nov 1; Jakarta, Indonesia. Jakarta: Politeknik Statistika STIS; 2021. p.1001-1011. doi:10.34123/semnasoffstat.v2021i1.889
7. Wijayanto AW, Triscowati DW, Marsuhandi AH. Maize field area detection in East Java, Indonesia: an integrated multispectral remote sensing and machine learning approach. In: *Proc 2020 12th Int Conf Inf Technol Electr Eng (ICITEE)*; 2020 Oct 6-8; Yogyakarta, Indonesia. Piscataway, NJ: IEEE; 2020. p.168-173. doi:10.1109/ICITEE49829.2020.9271683
8. United Nations Statistics Division (UNSD). Implementation guidelines: United Nations Fundamental Principles of Official Statistics [Internet]. New York: United Nations; 2015 [cited 2020 Mar 18]. Available from: [https://unstats.un.org/unsd/dnss/gp/Implementation\\_Guidelines\\_FINAL\\_without\\_edit.pdf](https://unstats.un.org/unsd/dnss/gp/Implementation_Guidelines_FINAL_without_edit.pdf)
9. United Nations Committee of Experts on Big Data and Data Science for Official Statistics. Geospatial data integration in official statistics. New York: United Nations; 2019.
10. United Nations Global Working Group on Big Data for Official Statistics. Satellite imagery and geospatial data task team report: integration of Earth observation data into official statistics. New York: United Nations; 2018.
11. Hassani H, Silva ES. Integrating geospatial and non-geospatial data for statistical analysis and decision-making. *Stats*. 2025;5(4):166.
12. United Nations Committee of Experts on Big Data and Data Science for Official Statistics. *Geospatial data integration in official statistics*. New York: United Nations; 2019.